



New England Ensembles

David J. Stensrud and Nusrat Yussouf
with help from many others!

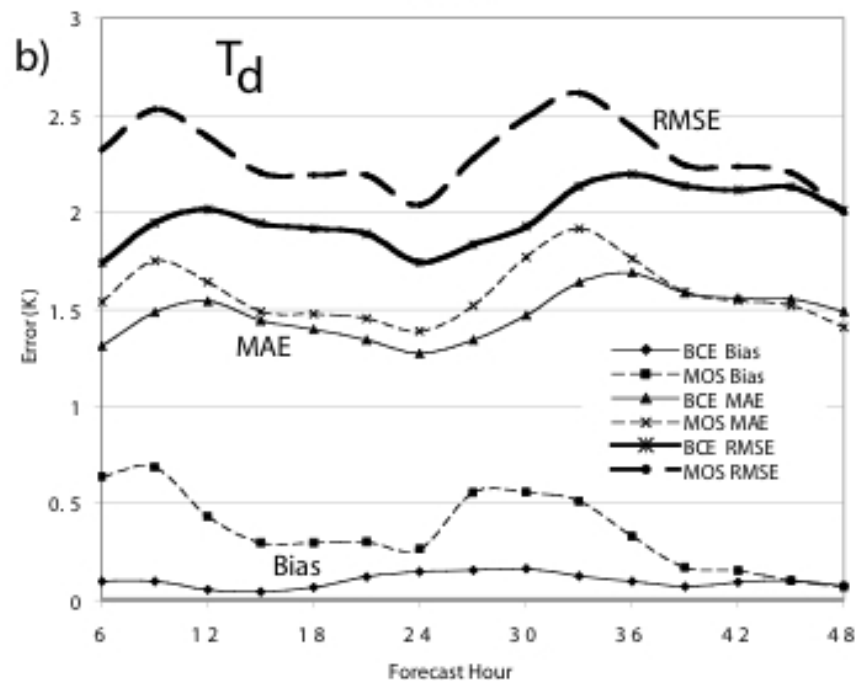
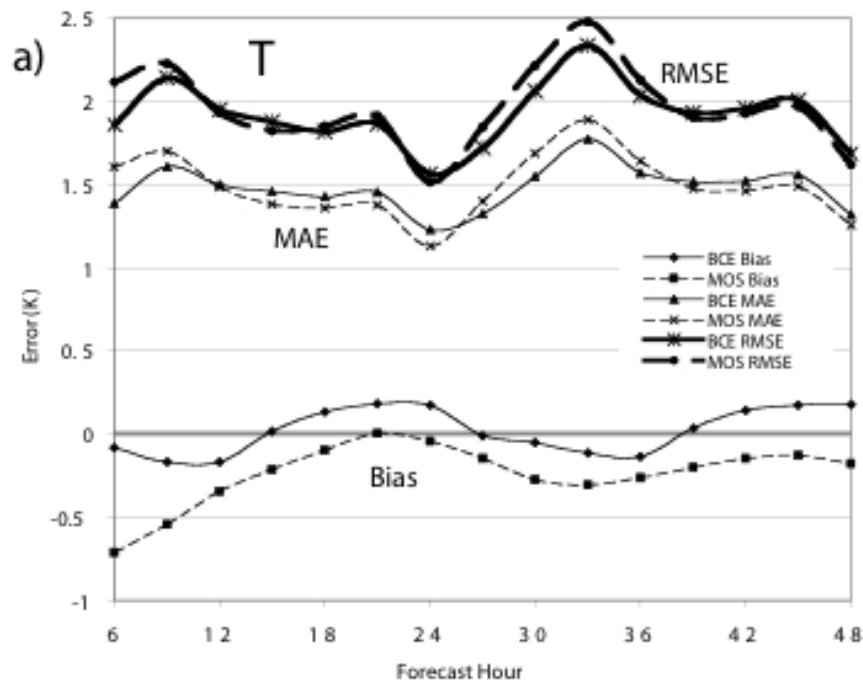
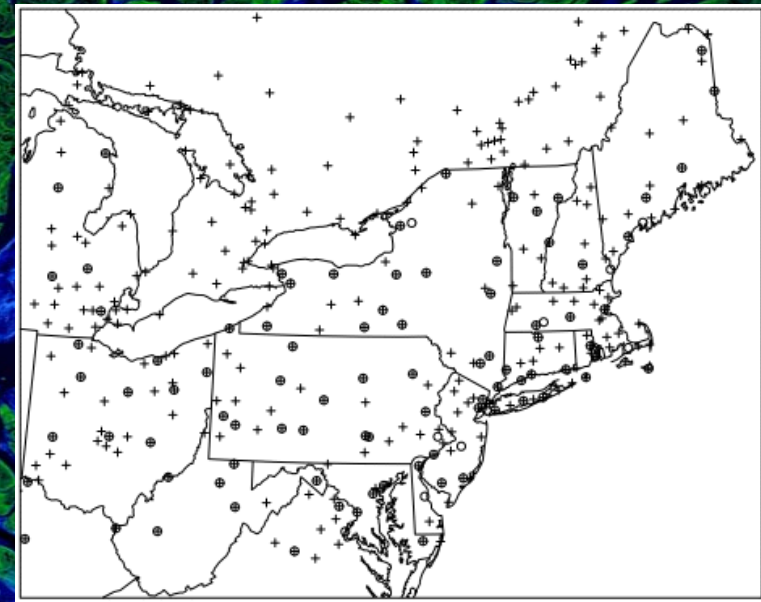


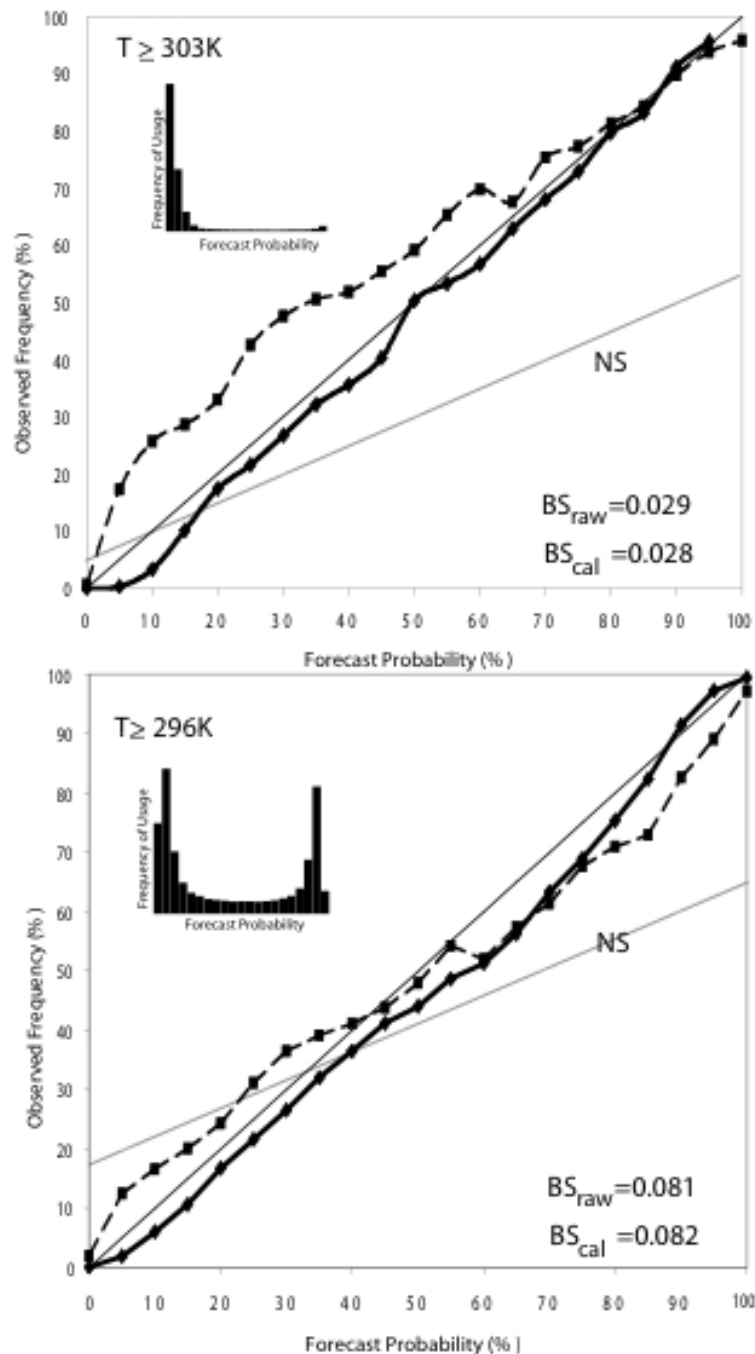
2002

- A little rushed
- Selected a 7-day bias correction window using the educated guess technique
- Collected a decent archive of 48 forecast days
- Put output on web in real time!
- All model data interpolated to a common 10-km grid for processing (made my life much easier!)

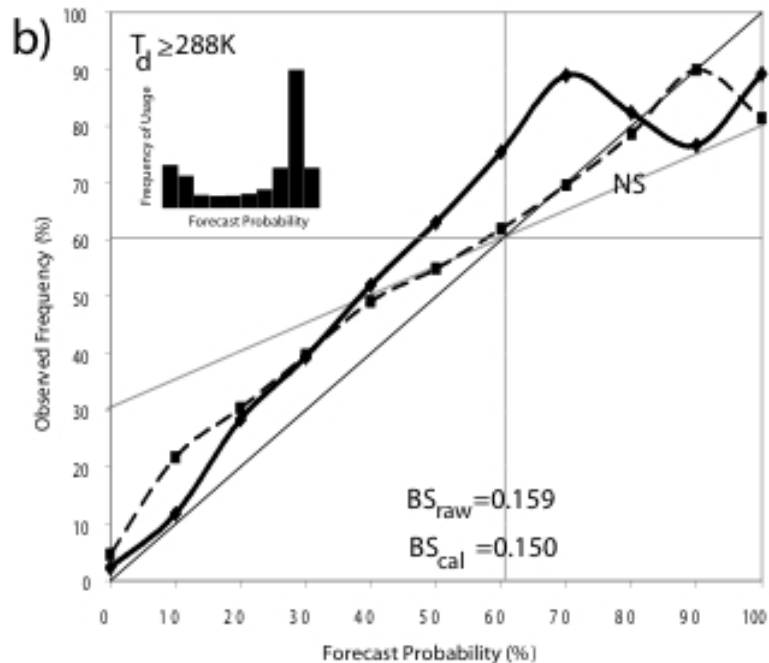
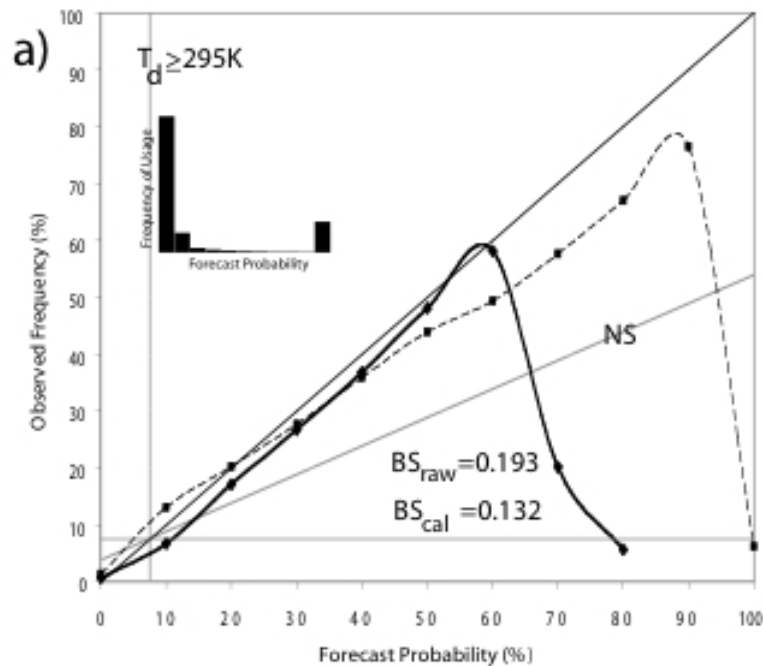
BCE mean comparable with NGM
MOS for T and BCE better
than NGM MOS for Td!

Not bad for using an
educated guess for the
window length!





- Results from 2002 also indicate that the BCE temperature probabilities are not too bad, and can be corrected using past 7-days of data to give reliable predictions

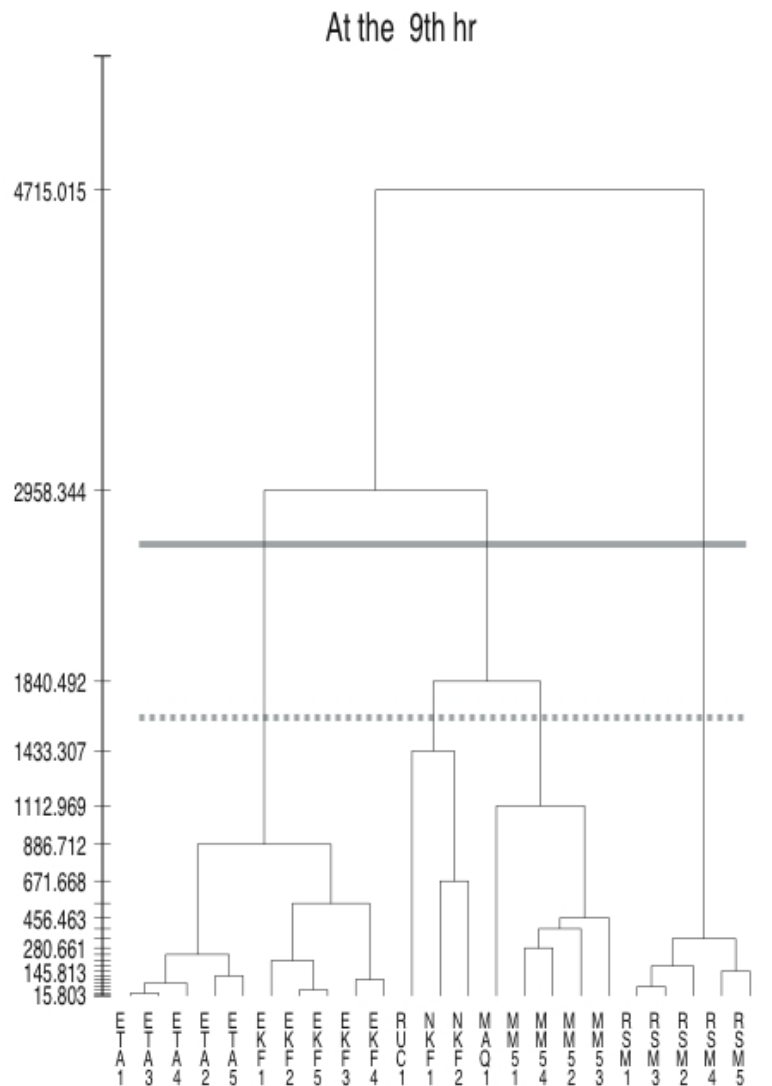


- Dewpoint temperature is a tougher nut to crack, and the results are not nearly as impressive



Cluster Analysis

- Used a cluster analysis approach to explore importance of both model and model physics diversity in the ensemble data using 23 forecast days for which all data are available at 2-m, 850 hPa, 500 hPa, and 250 hPa.
- Cluster analysis basically separates data into various groupings, and for our data these groupings match those determined by me for a handful of cases.



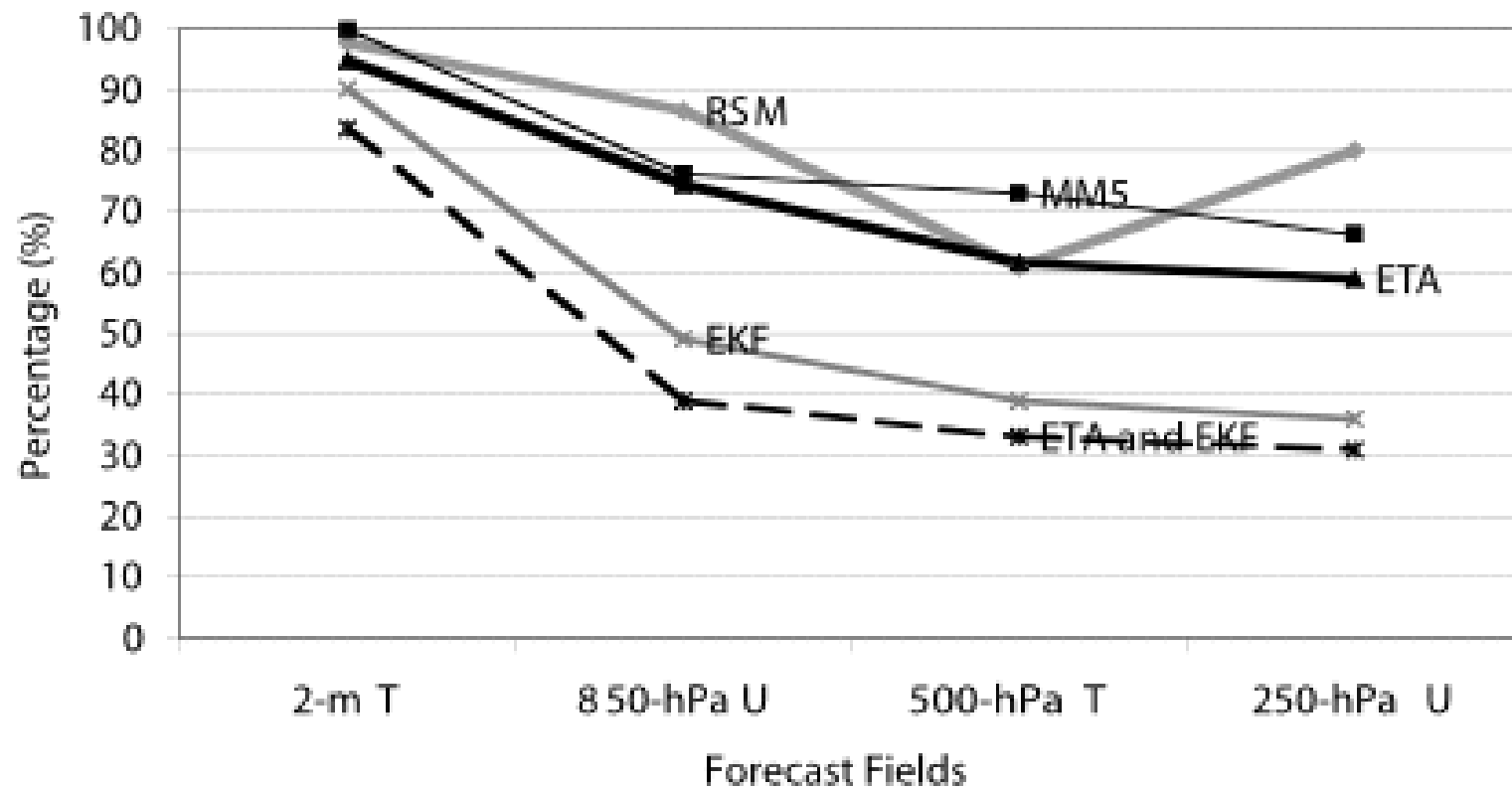
(a)

Dendrograms can be your friend!

Noticed that ETA1 and EKF1 have nearly identical initial conditions.

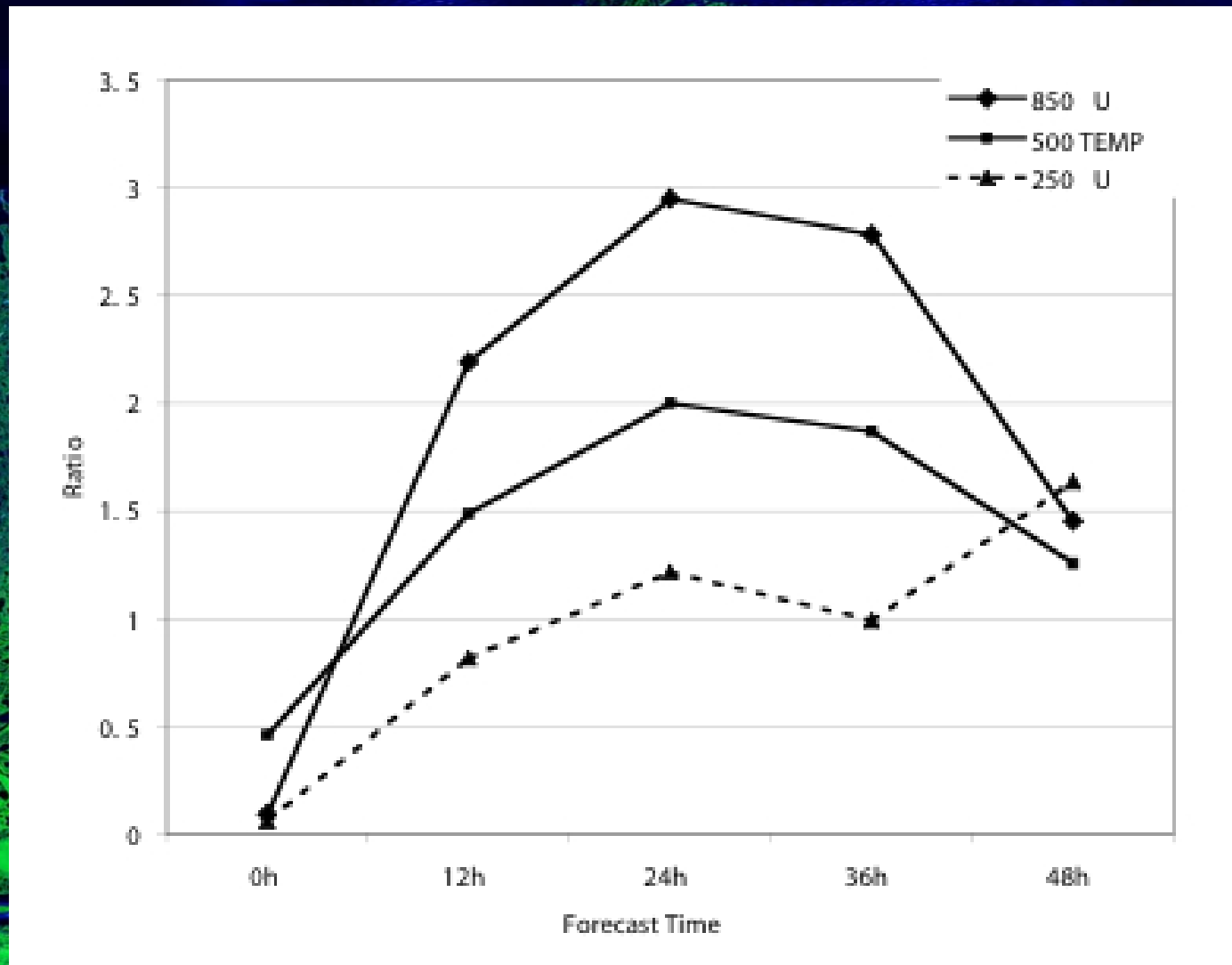
Explore how these two forecasts evolve.

← 13 / miles →



Results indicate that the models used often cluster first by model, especially closer to the surface.

→ 13 / miles →



Results show that EKF1 quickly moves outside of envelope of ETA ensemble members, indicating importance of model physics diversity to SREF.



2002 Conclusions

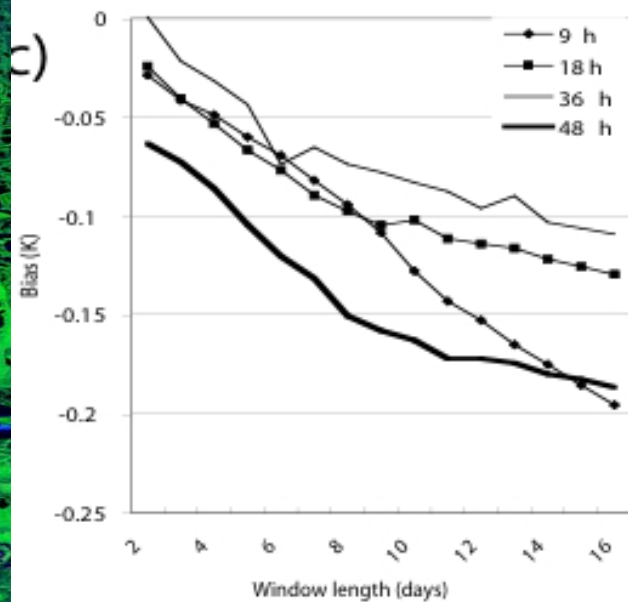
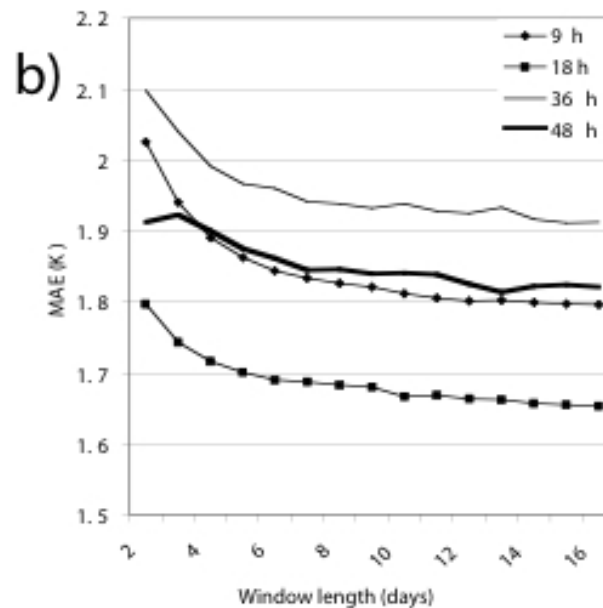
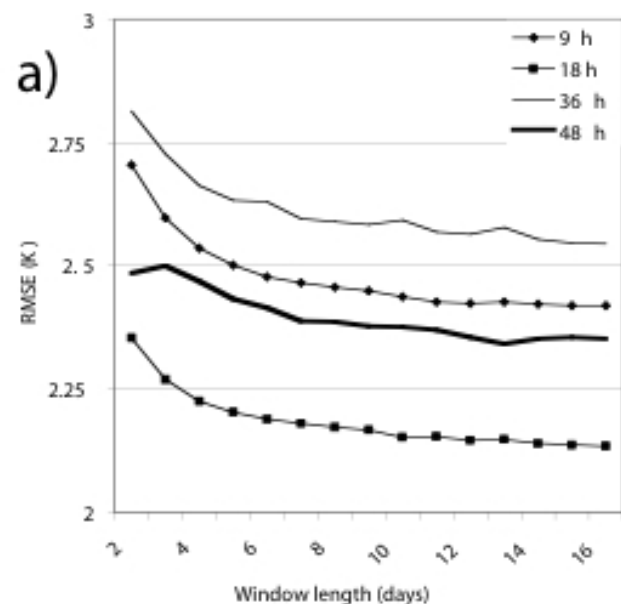
- BCE compares favorably against NGM MOS and better for dewpoint temperature
- Ways developed to provide reliable probabilities given threshold values, and cost-loss problems also indicate that the BCE is more valuable than NGM MOS
- Have no idea if bias correction window length is reasonable
- Cluster results clearly show that model and model physics diversity are important for SREF and that more diversity is needed near the land surface!



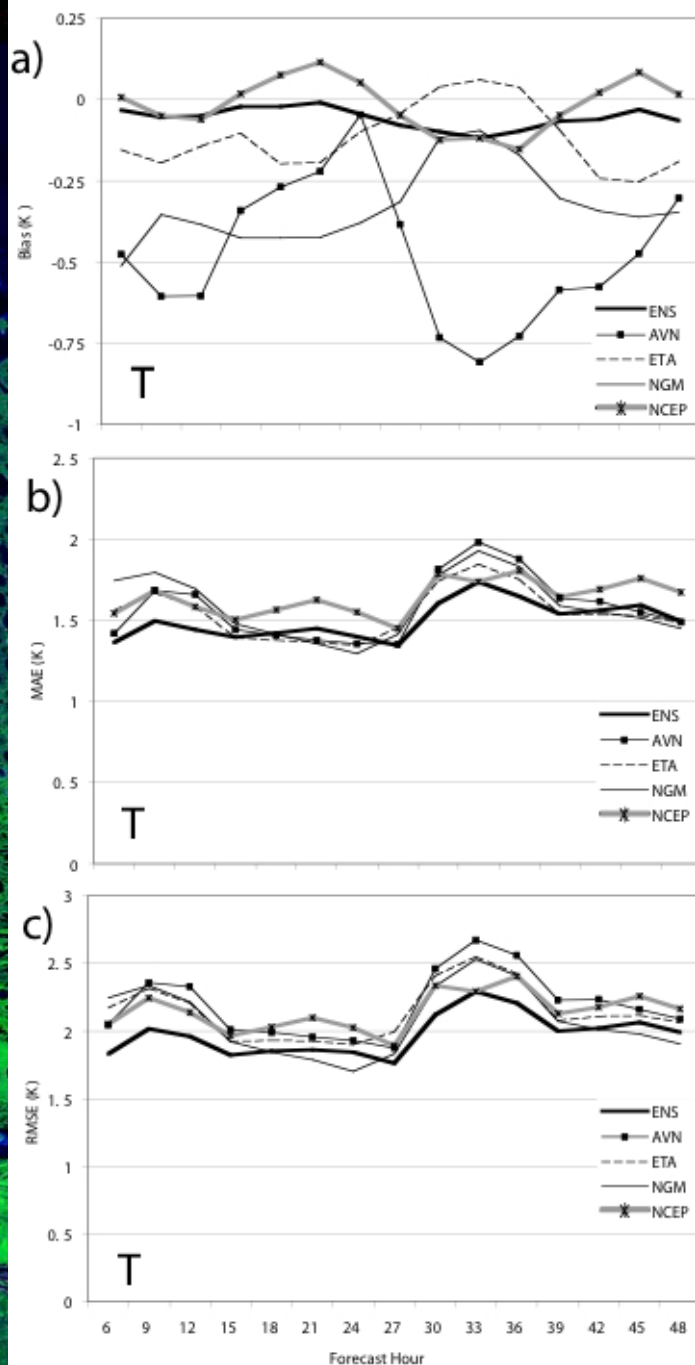
2003

- Not quite so rushed!
- Everyone had problems
 - RUC upgrade during early July
 - NSSL RAID failure during early July
 - NCEP SREF problems in late July/early August
- Yet still managed to collect 43 case days!
- Put output on web in real time with addition of histogram showing more information on distribution of forecast values
- Model data now collected in GRIB on native grid

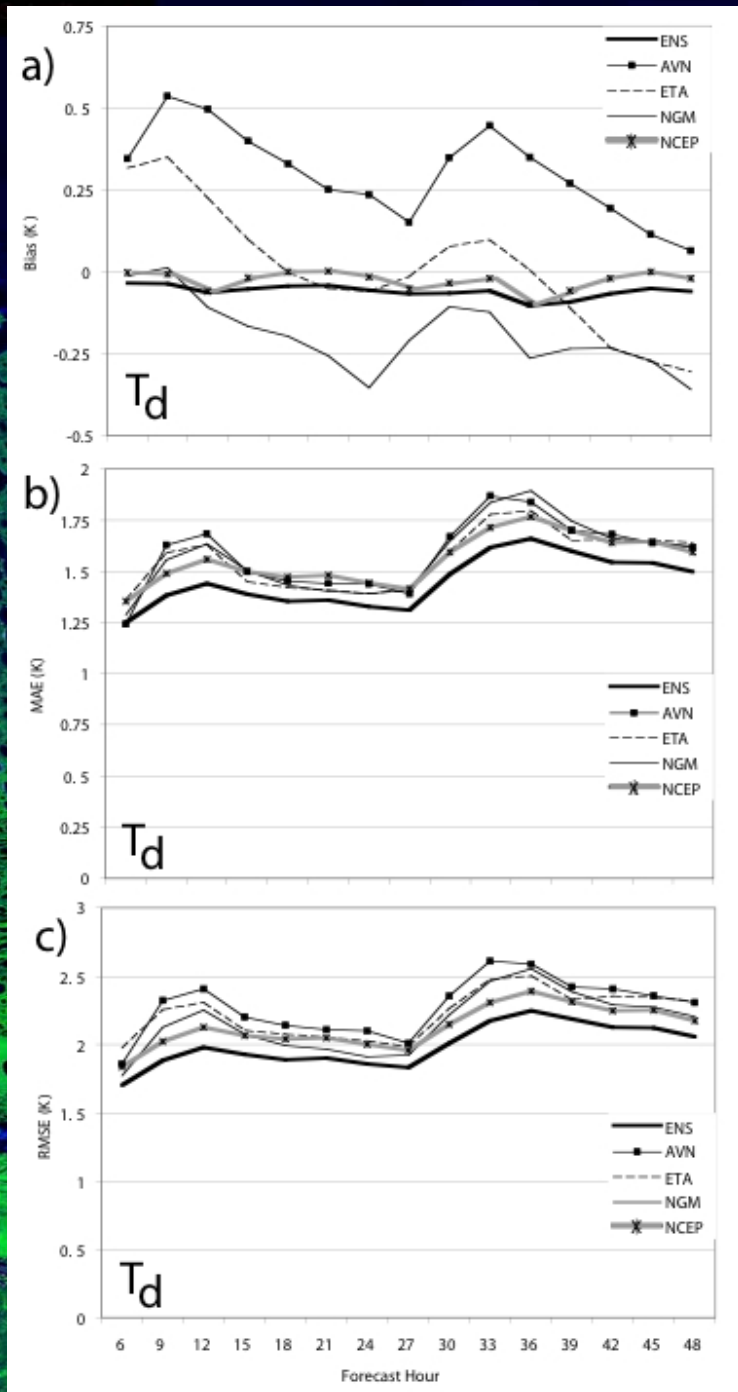
→ 13 ° miles →



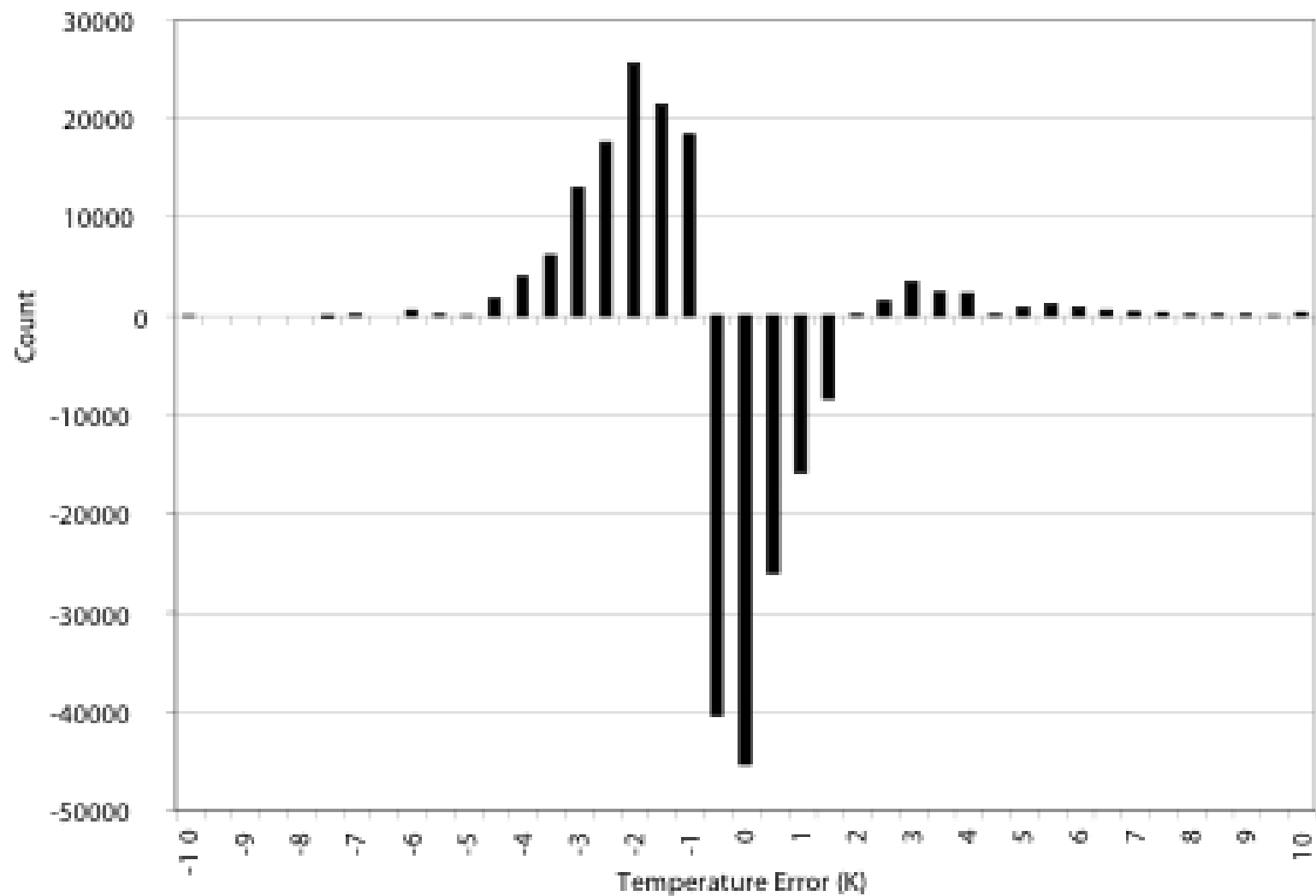
Bias correction
window length
of 12 days
seems to work
best, so while
7 days was a
reasonable
guess we can
do better!



- Mean BCE temperatures better than NGM MOS, Eta MOS, and AVN MOS during afternoon hours. These differences are significant at the 95% level or greater. Mean BCE based upon NCEP-only data comparable with MOS but not better than MOS.

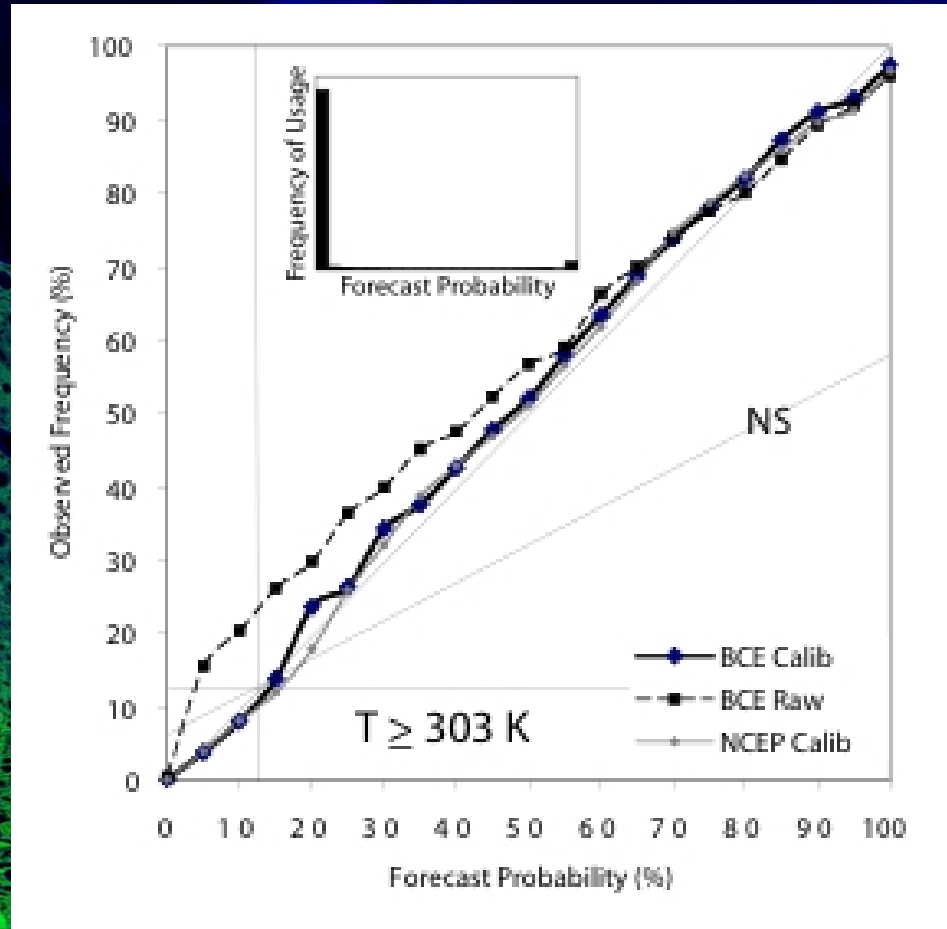


- Mean BCE dewpoint temperature is better than all the MOS products at nearly all forecast times! These differences also are significant at the 95% level or better. NCEP-only BCE again is comparable with MOS, but not better than.
- Results further show that mean BCE comparable with consensus MOS for T and T_d . Lower mean error of BCE when compared with consensus MOS, but lower rmse for consensus MOS when compared with BCE.



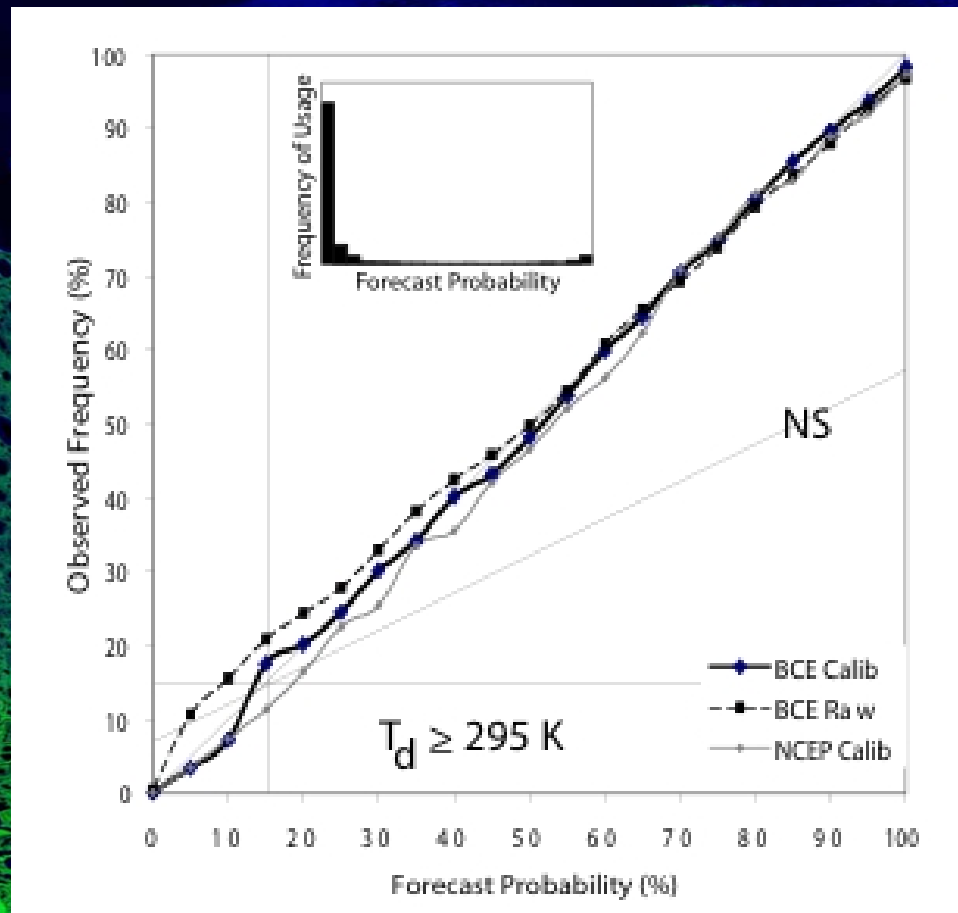
Mean BCE temperature forecasts have fewer large errors when compared against Eta MOS.

13 miles



- BCE temperature probability forecasts are better than in 2002 (raw values) and with correction are very reliable! NCEP-only BCE values nearly as good, and differences not significant.

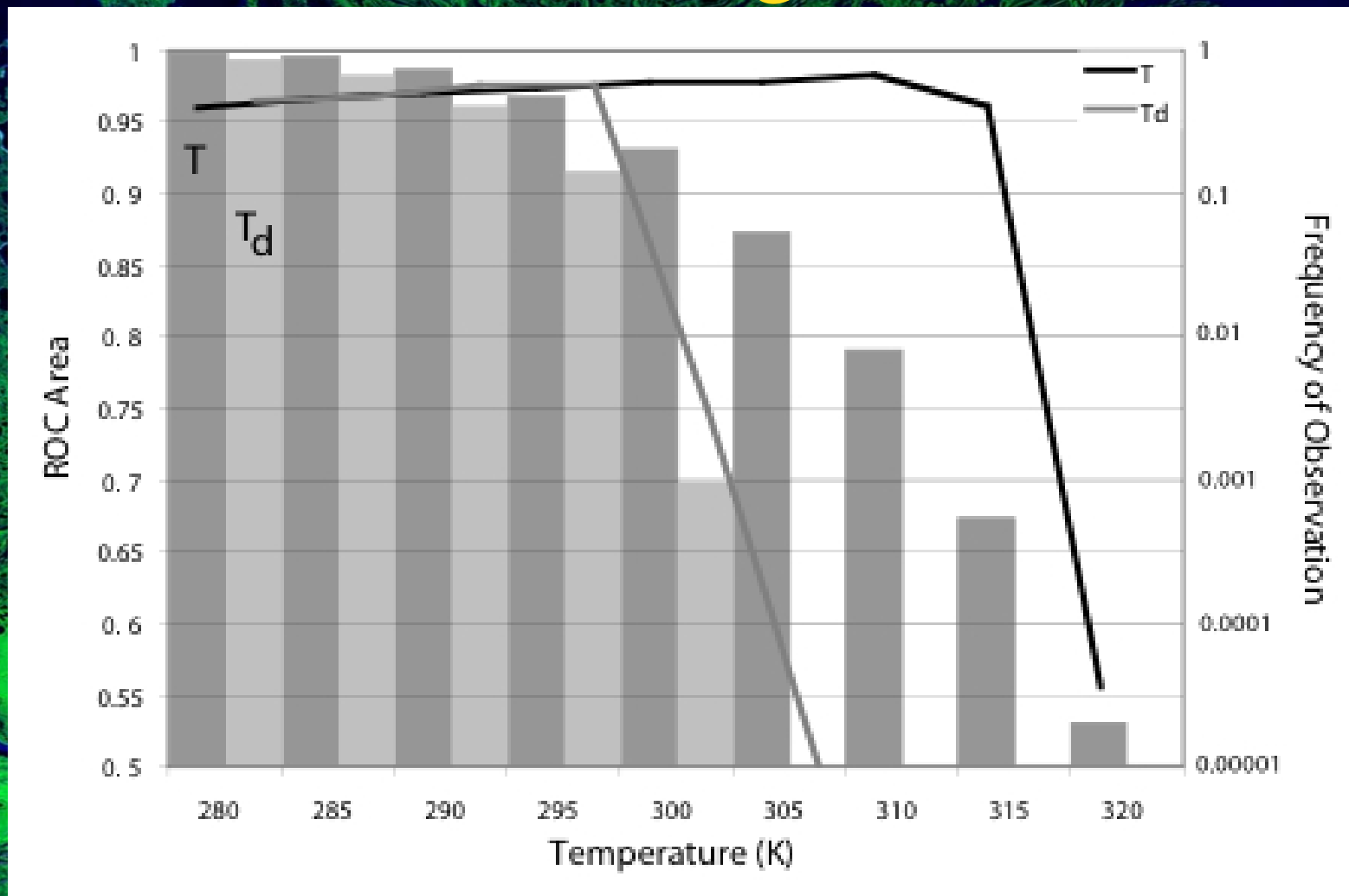
← 13 / miles →

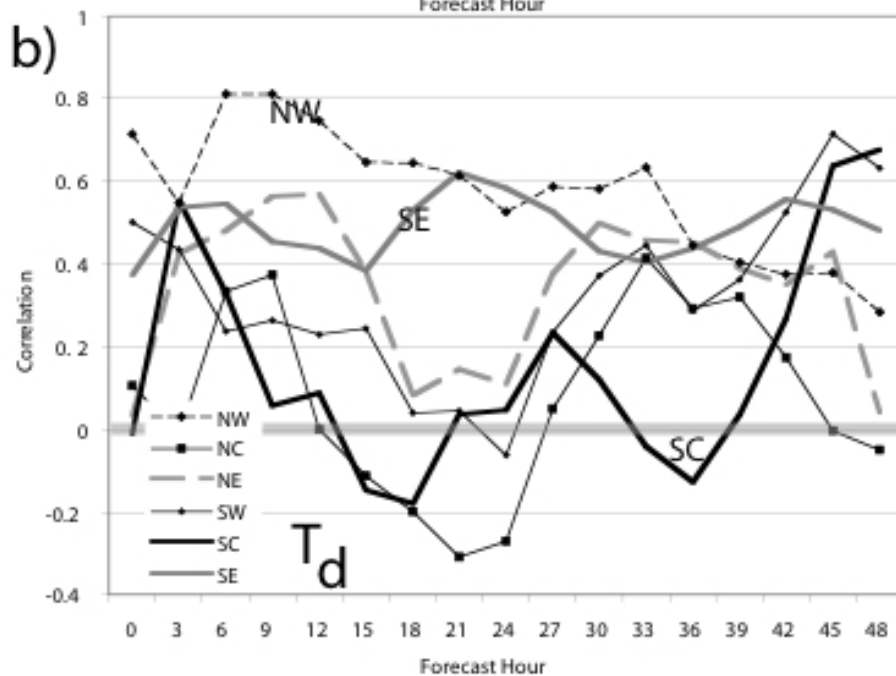
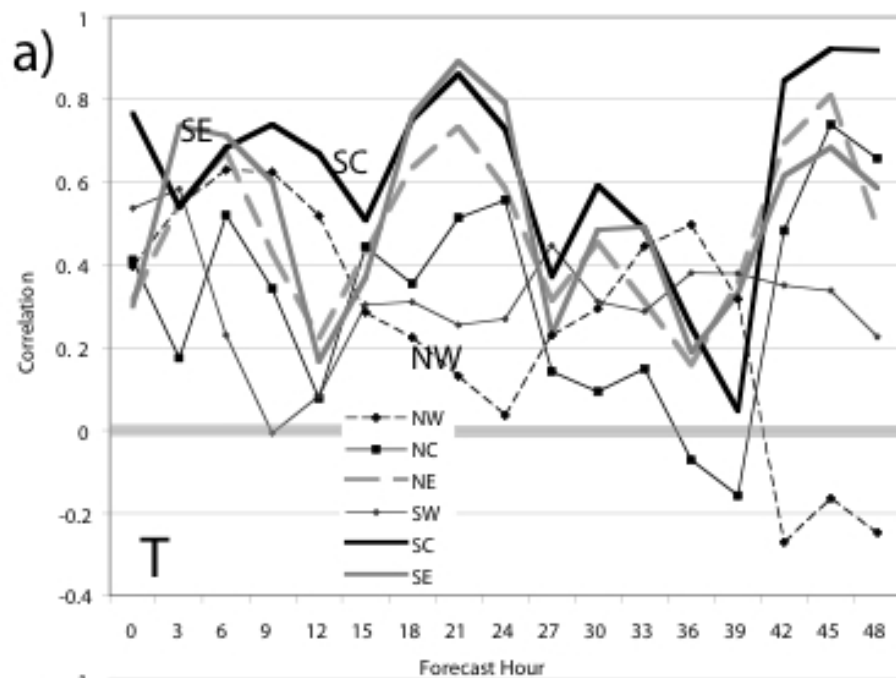


- BCE dewpoint temperature probability forecasts are MUCH better than in 2002 and with correction are very reliable. NCEP-only BCE values nearly as good, and differences not significant.

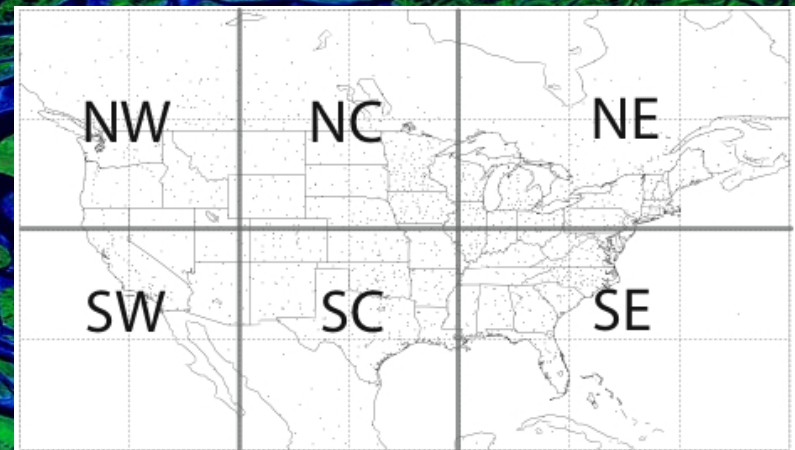
← 13 / miles →

ROC Diagrams

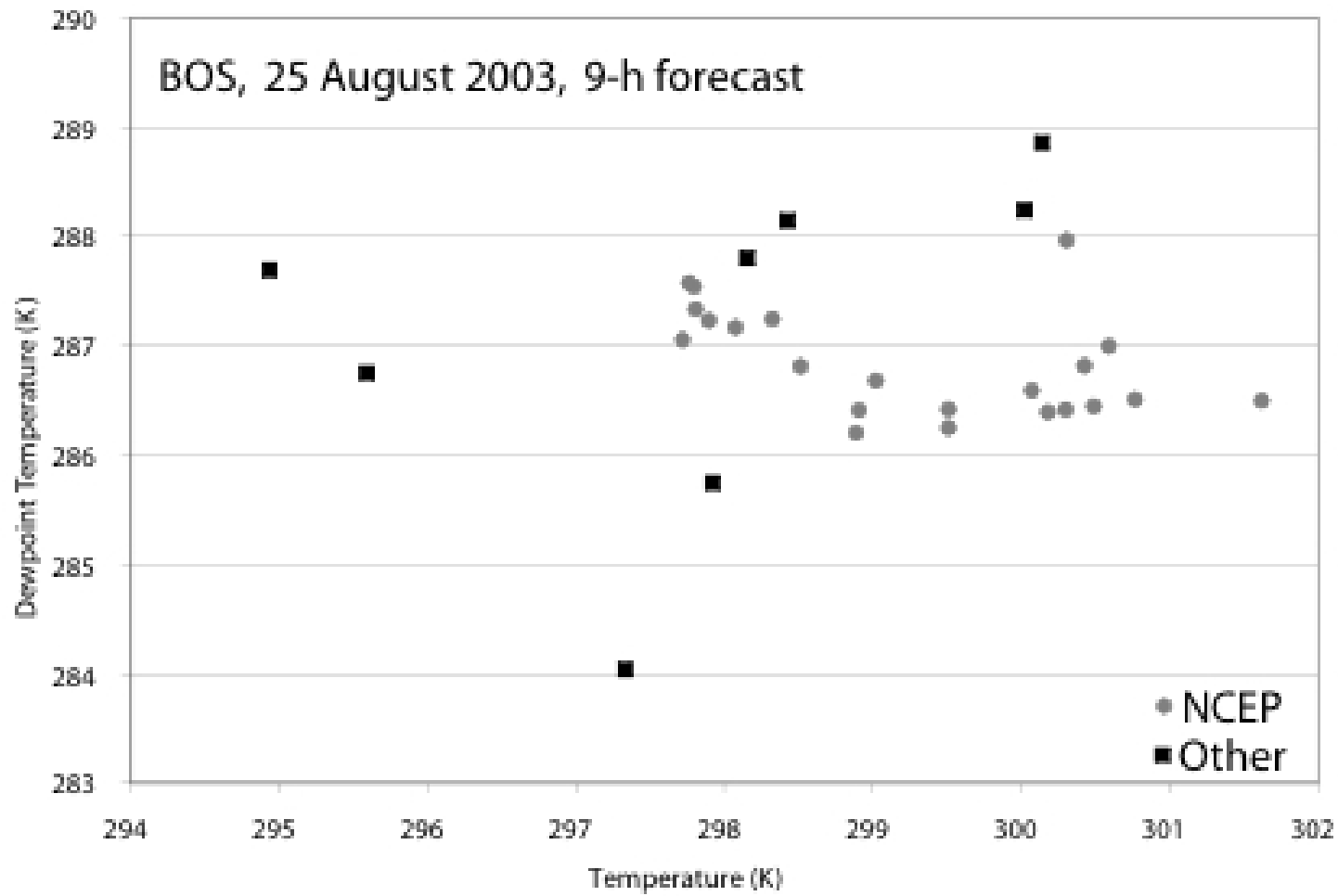




Spread-skill relationship is pretty good in some parts of the US for specific forecast times, but varies tremendously! A lot of curious behavior in these data!



What do extra members add?



2003 Conclusions

- Mean BCE T and Td forecasts generally better than all MOS products during the daytime, and comparable with consensus MOS.
- BCE probabilities more reliable than in 2002, especially for dewpoint.
- ROC diagrams indicate skill in BCE probabilities for wide range of observed values.
- Spread-skill relationship still apparent, and now extends to dewpoint temperature as well, for certain parts of the US.
- Results of wind speed calculations just being completed.